



TREBALL FINAL DE GRAU



ESCOLA
POLITÀCNICA SUPERIOR
UNIVERSITAT DE LLEIDA
INSPIRING THE FUTURE

Estudiant: Arnau Benavides Roca

Titulació: Grau en Enginyeria Informàtica

Títol de Treball Final de Grau: **¿Quina edat pulmonar tinc?**

Director/a: **Jordi Vilaplana Mayoral, Marc Pifarré Montalà**

Presentació

Mes: Juny

Any: 2021

Resum

Fonaments: Actualment, les proves d'espirometria no són accessibles a tothom, ja sigui pel temps d'espera, el cost en una clínica privada o la ubicació del pacient entre d'altres.

Objectius: L'objectiu d'aquest treball és poder obtenir l'edat/capacitat pulmonar a través de la mostra d'àudio d'una bufada.

Mètodes: Realització d'una aplicació web per obtenir mostres, scripts per detectar la bufada i extreure'n les característiques, analitzar els millors resultats de les tècniques de machine-learning/intel·ligència artificial.

Resultats: Tot i no ser els que esperàvem no són dolents, ja que podem arribar a predir més de la meitat de mostres.

Conclusions: Generalment ens han faltat mostres per poder obtenir millors resultats en la predicció.

Paraules clau: Espiròmetre, app, exhalar, predicció de la capacitat pulmonar, machine learning, intel·ligència artificial

1. Introducció

Les malalties respiratòries causen una gran càrrega sanitària, econòmica i social. Aquestes són la tercera causa de mort del món [1]. Per això, s'han dedicat molts esforços a l'investigació del diagnòstic prematur i al seguiment rutinari dels pacients amb malalties respiratòries per poder intervenir a temps [2].

L'alè exhalat pels humans està format per aerosols, composts orgànics volàtils i no volàtils, fosfolípids [3], proteïnes, productes d'oxidació i microbiomes [4, 5, 6]. Les diferents condicions patològiques provoquen que les composicions d'aquestes barreges siguin diferents, això condiciona al fet que tinguin un patró únic ('empremta digital') específic per cada tipus de malaltia. Tanmateix, el que ens interessa és un altre tipus d'empremta, la produïda en exhalar. La idea és utilitzar-la per predir la capacitat pulmonar tal com faria un espiròmetre.

Durant una prova d'espirometria el pacient agafa aire de la manera més profunda i possible que pugui. Llavors exhala amb la màxima força durant el màxim de temps possible. L'espiròmetre mesura la quantitat i la velocitat del flux d'aire i calcula diverses funcions pulmonars basant-se en la prova [7].

Les quatre mesures més importants de la funció pulmonar són: (Figura 1):

1. Capacitat Vital Forçada (Forced Vital Capacity, FVC): volum total d'aire expulsat durant l'expiració.
2. Volum espiratori forçat (Forced Expiratory Volume) en un segon (FEV1): volum d'aire expulsat en el primer segon de l'expiració.
3. FEV1/FVC (FEV1%): relació entre FEV1 i FVC.
4. Flux màxim d'expiració (Peak Expiratory Flow, PEF): cabal d'aire màxim assolit durant l'expiració.

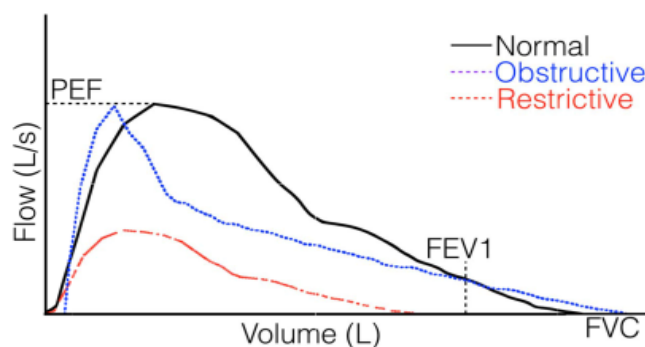


Figura 1: Espirometria. Font: [7].

Les mesures de la funció pulmonar d'una persona sana solen ser d'almenys el 80% dels valors predits en funció de la seva edat, estatura i sexe [8]. Els valors anormals de FEV1% són (expressats com el percentatge del valor predit) [9]:

- Disfunció pulmonar lleu: 60-79%
- Disfunció pulmonar moderada: 40-59%
- Disfunció pulmonar greu: inferior a 40%.

El gràfic de la FVC és similar a una funció de densitat exponencial (línia sòlida i negra a la Figura 1) a persones sanes. A mesura que augmenta l'obstrucció al flux d'aire, la taxa de flux disminueix més ràpidament que de forma exponencial després d'arribar al seu valor màxim (PEF). Per tant, aconsegueix una distribució esbiaixada (línia discontinua blava a la Figura 1). Quan es pateix una malaltia pulmonar restrictiva, com la fibrosi quística, els músculs respiratoris es debiliten i la capacitat pulmonar del pacient (FVC) disminueix (línia discontinua vermella a Figura 1). La forma és molt similar a una distribució de Weibull.

A continuació, s'extrauran les característiques estimades FVC, FVC1% i PEF. Per exemple, la característica estimada PEF serà la freqüència màxima obtinguda a la representació de temps-freqüència de la mostra. La característica d'estimació del FVC serà l'energia total del gràfic temps-freqüència. La característica FEV1% és la relació entre FEV1 estimada i FVC.

Segons la Figura 1 podrem diagnosticar al pacient amb una disfunció pulmonar. No ens interessa diagnosticar una malaltia pulmonar determinada, sinó només si la persona presenta algun tipus de símptoma d'alguna malaltia pulmonar.

Mentre una persona bufa, les cordes vocals està inactives. Només la posició de la boca afecta la naturalesa del so emès [10]. L'emissió d'un soroll de bufar pot ser aproximada per una font de soroll blanc que passa per un filtre de banda. Es consideren dos tipus de soroll de bufada, denominats bufada calenta i freda. Un soroll de bufada calenta és el tipus de soroll produït per algú que intenta entelar el vidre d'una finestra. Un soroll de bufada freda, per exemple és el soroll d'algú que està refredant un plat de sopa. La Figura 2 mostra l'espectrograma de la potència d'aquests dos tipus de soroll de bufada, així com el de parlar per comparar.

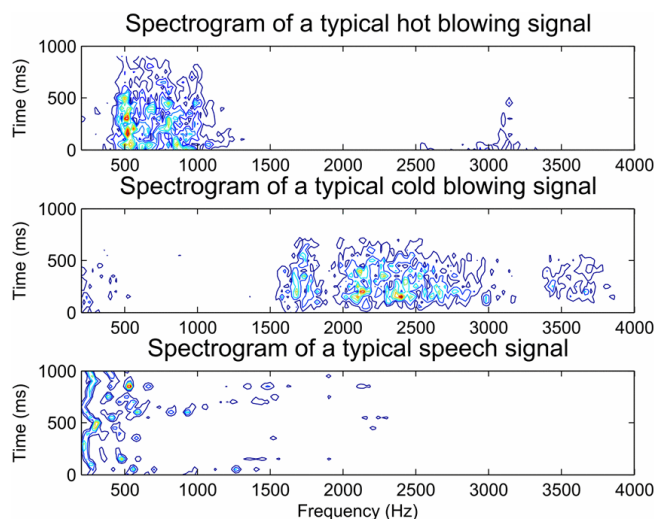


Figura 2: Espectrograma. Font: [10].

Pels dos tipus de soroll bufat, l'espectre no és periòdic i la seva energia està clarament localitzada en diferents parts de l'espectre. El senyal de la parla conté components periòdiques amb freqüències fonamentals compreses entre 85 i 255 Hz [11].

A la mateixa mostra obtinguda en l'espirometria, s'obtindrà el patró de freqüència de l'espectrograma. S'espera que les bufades normals tinguin un patró de freqüència freda.

El principal problema actualment és que una espirometria s'ha de fer a un hospital o un centre mèdic on disposin d'un espiròmetre. Això pot arribar a tenir un gran temps d'espera i poder costos addicionals per al pacient en cas d'anar a una clínica privada. En aquest cas SpiroSmart ha demostrat que és possible realitzar una prova d'espirometria utilitzant un telèfon intel·ligent.

L'objectiu d'aquest estudi és emular una espirometria utilitzant el micròfon del telèfon mòbil perquè aquest procés sigui més fàcil i econòmic que la prova d'espirometria de l'hospital. Alhora que també sigui accessible per tothom.

2. Mètodes

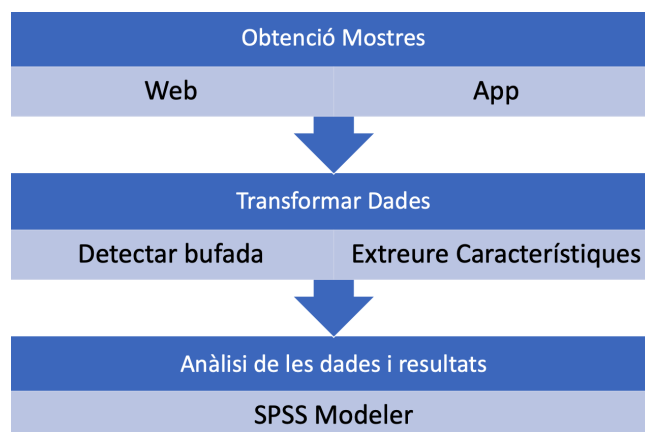


Figura 3: Procediments generals dels mètodes

Com podem observar a la Figura 3 s'han seguit un conjunt de passos per poder dur a terme aquest projecte i que més endavant s'explicaran detalladament. En primer lloc tenim l'obtenció de mostres, que per aconseguir mostres s'ha desenvolupat una aplicació web i una aplicació mòbil i se n'ha fet difusió per arribar a tenir el màxim de mostres possibles. Un cop es tenen mostres les hem de transformar, això significa eliminar mostres no vàlides o incorrectes i extreure característiques per un posterior anàlisi que el farem amb SPSS Modeler, una eina molt potent pel tractament de dades i l'anàlisi.

2.1. Obtenció de mostres

Una de les fases més importants del projecte és l'obtenció del conjunt de dades. La precisió dels classificadors depèn en gran part de la quantitat i la qualitat de les mostres que es tingui. Hem desenvolupat una aplicació web i una aplicació mòbil. Tot i que l'aplicació web ha sigut on hem centrat més esforços, ja que és on hem obtingut quasi totes les mostres. Això es deu al fet que els usuaris prefereixen accedir a una web momentàniament que no pas instal·lar una aplicació als seus dispositius mòbils. Aquestes aplicacions ens serveixen per capturar les mostres i les dades dels usuaris per posteriorment processar-les.

2.1.1. Aplicació Mòbil

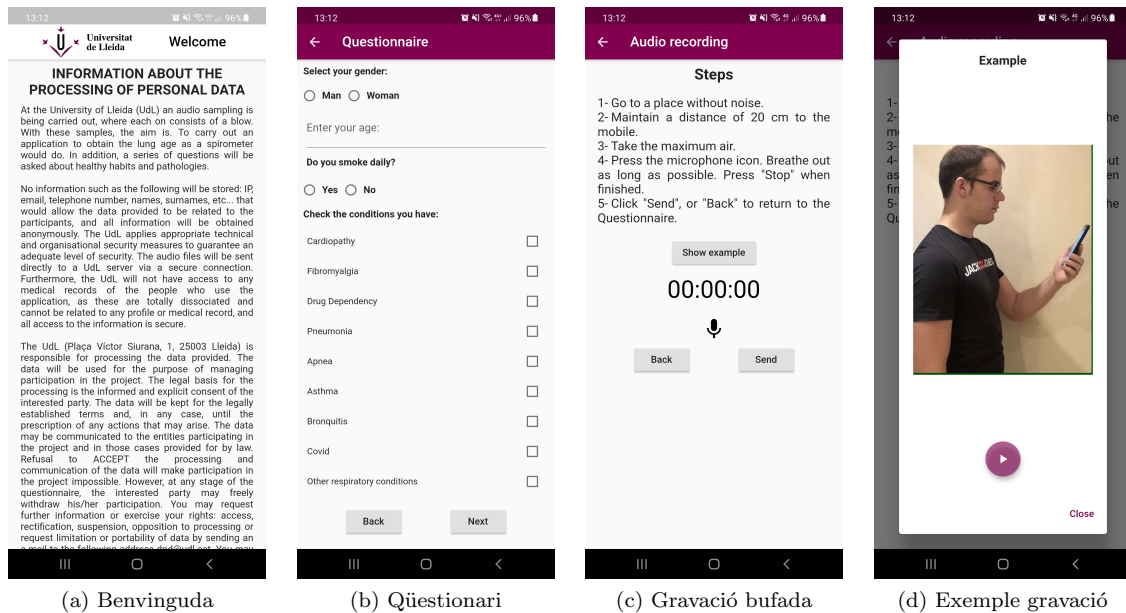


Figura 4: Aplicació mòbil per recollir dades.

En primer lloc, Fig. 4a mostra una pantalla de benvinguda per acceptar la política de privacitat per la recollida de dades en entrar a l'aplicació. Un cop acceptada, es mostra un qüestionari Fig. 4b. Per estimular l'enquesta i eliminar les barreres que disminueixen el seu èxit (molta diversitat), s'han unit els símptomes que poden influenciar en la tos i respiració en 7 afeccions principals: cardiopaties, fibromiàlgia, pneumònia, apnea, asma, bronquitis i tabaquisme. En la figura Fig. 4c es mostra la pantalla de l'aplicació pel registre de mostres d'exhalació on s'explica el protocol a seguir. Aquest protocol intenta ser senzill i comú per facilitar l'ús als usuaris un cop l'espírometre estigui finalment desenvolupat. Consisteix a mantenir una distància de 20 cm aproximadament entre la boca i el telèfon mòbil. A continuació, agafar el màxim d'aire possible i bufar el màxim d'aire possible. També s'ha proporcionat un vídeo d'ajuda on es pot veure aquest procés Fig. 4d.

2.1.2. Aplicació Web

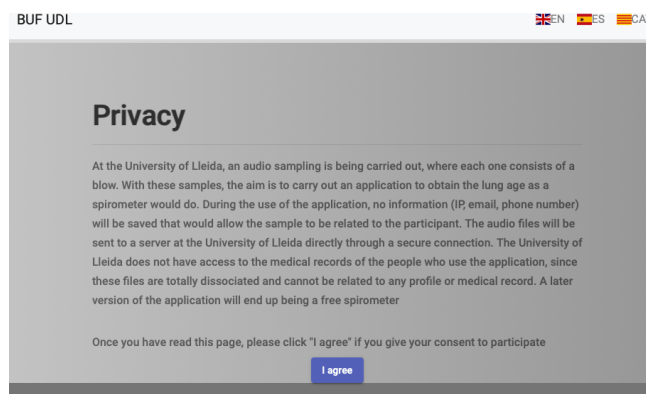
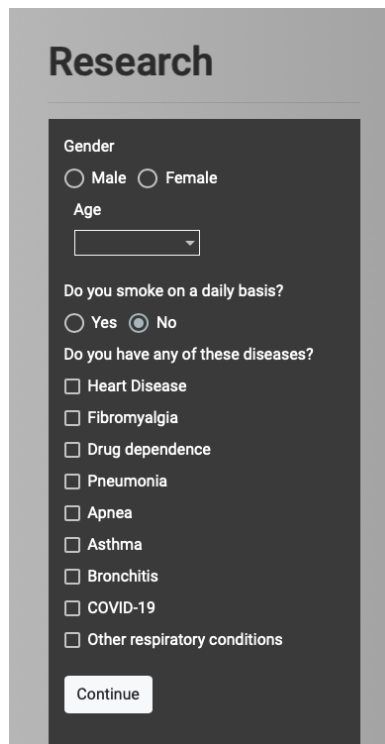


Figura 5: Benvinguda.



Research

Gender
☐ Male ☐ Female

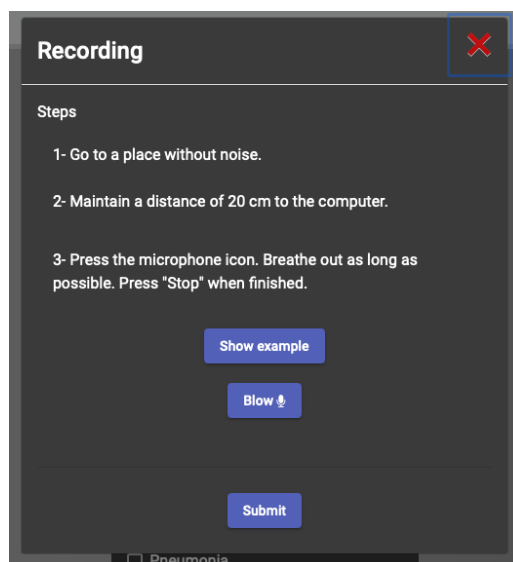
Age

Do you smoke on a daily basis?
☐ Yes ☒ No

Do you have any of these diseases?

- ☐ Heart Disease
- ☐ Fibromyalgia
- ☐ Drug dependence
- ☐ Pneumonia
- ☐ Apnea
- ☐ Asthma
- ☐ Bronchitis
- ☐ COVID-19
- ☐ Other respiratory conditions

Figura 6: Questionari.



Recording ✕

Steps

- 1- Go to a place without noise.
- 2- Maintain a distance of 20 cm to the computer.
- 3- Press the microphone icon. Breathe out as long as possible. Press "Stop" when finished.

☐ Pneumonia

Figura 7: Gravació bufada.

Login

Username *

Password *

Login

Figura 8: Inici sessió administrador.

BUF UDL EN ES CAT

Download all audios Download info Summary

Filter
All

Gender	Age	Smoker	Conditions	Submitted	Blow	Status
M	26	No	Asthma	23/05/2021 20:58		New Save Delete
M	46	No	Other	23/05/2021 20:58		New Save Delete
M	32	Yes		23/05/2021 20:58		New Save Delete
M	69	No		23/05/2021 20:58		New Save Delete
M	39	Yes	Apnea	23/05/2021 20:59		New Save Delete
M	37	No		23/05/2021 20:59		New Save Delete
M	68	Yes		23/05/2021 20:59		New Save Delete
M	24	No		23/05/2021 21:00		New Save Delete
M	38	No		23/05/2021 21:00		New Save Delete
M	44	Yes	Heartdisease	23/05/2021 21:00		New Save Delete

Items per page: 10 1 - 10 of 29 |< < > >|

Figura 9: Inici administrador.

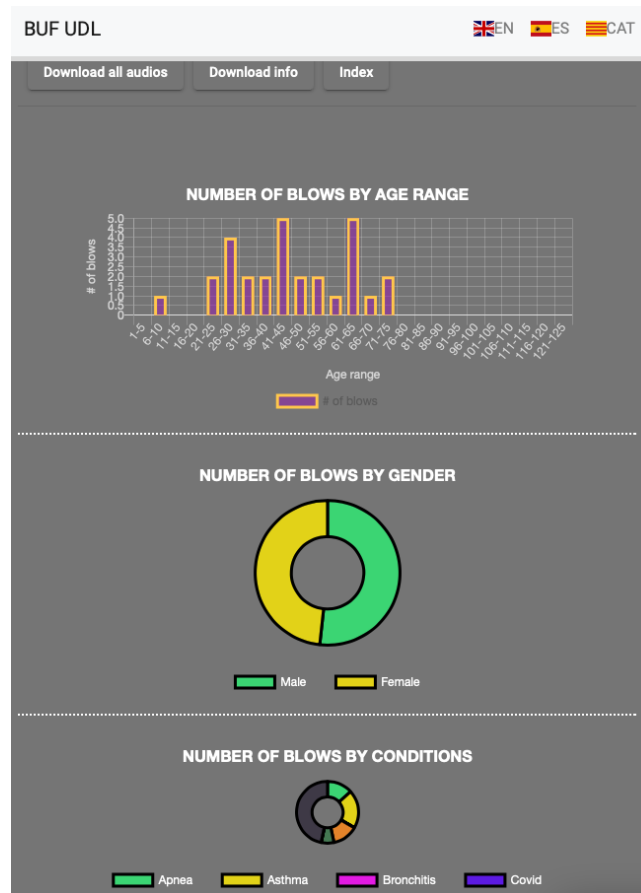


Figura 10: Resum dades.

En primer lloc, quan entrem a la pàgina web veiem Fig. 5 que és una pàgina de benvinguda per acceptar la política de privacitat per la recollida de dades en entrar a la pàgina web. Un cop donat el consentiment, es mostra un qüestionari Fig. 6. El qüestionari contindrà les mateixes qüestions que en l'aplicació mòbil, ja que finalment les dades de les dues aplicacions han de ser les mateixes. Quan s'ha introduït el qüestionari i es prem al botó de "Continuar" ens mostrarà una finestra emergent Fig. 7 on hi ha una breu explicació del protocol a seguir per enregistrar la mostra d'àudio. També podem visualitzar l'exemple de com seguir el protocol i enregistrar.

A la mateixa web, disposem d'una pàgina per administrar les dades. Perquè aquestes dades estiguin segures, la pàgina d'administrador està tancada i es necessita autenticar tal com podem observar a Fig. 8. Un cop passada l'autenticació veiem la pàgina inicial d'administrador Fig. ?? on destaquen 3 botons i una taula. La taula està composta en cada fila el qüestionari i la mostra d'àudio per cada usuari que l'ha realitzat. Podem escoltar i descarregar l'àudio de la fila corresponent que cliquem. Si es detecta que una mostra o enquesta és incorrecta es pot eliminar i així no interferir en l'anàlisi posterior.

La funcionalitat del botó superior de l'esquerra és de descarregar totes les mostres d'àudio en un fitxer comprimit, així es facilita la descàrrega i no fa falta descarregar un a un. El botó del mig ens descarregarà un fitxer en format '.csv' que contindrà totes les dades de les enquestes.

Finalment l'últim botó ens mostrarà la pàgina Fig. 10, que és un resum de les dades que hi han introduïdes. Per fer-ho més visual que en una taula, s'han dibuixat els gràfics perquè ràpidament es pot veure quina quantitat tens de cada mostra o si tens un dèficit de mostres en un rang d'edat.

Aquesta aplicació web ha estat desenvolupada amb Angular [12]. Angular és un framework opensource que facilita el desenvolupament tant del 'Frontend' com el 'Backend'.

Les dades un cop introduïdes a l'aplicació web, s'emmagatzema a una base de dades de mongoDB [13]. Així

quan s'han de mostrar l'únic que hem de fer és consultar aquesta base de dades. Això permet tenir un ordre i estructura ames que sigui escalable en un futur si fes falta.

2.2. Scripts detectació audio

Per extreure les característiques necessàries de les mostres d'àudio recollides, com els usuaris deixen espais sense bufar a la mostra s'ha programat un petit script per detectar exactament on està la bufada o fins i tot descartar l'àudio si no detecta correctament cap bufada. Quan s'han processat totes les mostres el següent pas és processar-les amb el Matlab [14] per extreure'n les característiques que després farem servir per analitzar i predir.

2.2.1. Detectar i tallar audio

La finalitat del script és descartar els àudios que no siguin correctes i detectar els silencis de l'inici i el final de la bufada, ja que si es processen amb aquests silencis les dades obtingudes, resultaran ser incorrectes perquè no només s'està agafant la bufada. Aquest script està programat amb python [15] i s'utilitza la llibreria PyDub [16] pel tractat d'àudio. Aquesta llibreria ens permet descompondre l'àudio en un array i així poder-lo manipular fàcilment. Per detectar els silencis, s'estableix un índex per considerar si és silenci i quan ja no ho és.

En una funció recursiva recorrem l'array des de l'inici mentre no se superi l'índex pel silenci, un cop se supera i es considera que ja és bufada ens n'assegurem que sigui la bufada d'una llargada mínima. En cas contrari es considerarà un fals soroll i es tornarà a repetir el procediment des del fals soroll fins que trobem la bufada. El resultat d'aquest procediment és detectar i tallar el principi de la mostra d'àudio. Per detectar i tallar el final de la mostra d'àudio, s'utilitza el mateix procediment però amb l'àudio girat. El resultat final és el fitxer d'àudio tallat amb la bufada, sense silencis ni sorolls.

2.2.2. Extreure característiques

	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA
1	H_f	Inf_Mut	fm	kurt_Mgt	MomC_11	MomC_77	MomC_1515	MomM_11	MomM_77	MomM_1511	sex	age	fuma	cardio	pneumo	drogodep	fibro	apnea	asma	bronq	others	max_db	med_db
2	9.42769988	0.36498881	1314.44526	1.60773786	-0.3030567	1.09639566	0.49684601	0.30693049	-0.1809	-0.168019	H	18	False	False	False	False	False	False	False	False	False	24.84044	-44.861725
3	9.15299551	0.21948042	1895.89013	1.6948316	-0.9170503	-0.2990201	-0.1095544	-0.2229674	-0.2581443	-0.1703218	H	18	False	False	False	False	False	False	False	False	False	-0.7559849	-69.926796
4	9.20450055	-0.05312	1647.91586	2.04049242	-0.0572124	-0.0233636	-0.0238193	0.44731816	0.18978976	0.03826318	H	19	False	False	False	False	False	False	False	False	False	41.48546	-18.433
5	9.19263898	-0.1519462	1814.3144	1.70766414	0.07656398	0.10523795	0.09678786	-0.6435242	-0.5180109	-0.3721852	H	20	False	False	False	False	False	False	False	False	False	39.215683	-40.784317
6	9.2017962	-0.495813	1708.92228	1.77189678	0.1845952	0.0485082	-0.0037758	0.73430329	0.51355186	0.27271501	H	20	False	False	False	False	False	False	False	False	False	30.09937	-48.054935
7	9.18683659	-0.1057053	1745.05204	2.26715533	0.06106881	-0.0715204	-0.1016699	-0.6788573	-0.2930365	-0.1777789	H	21	False	False	False	False	False	False	False	False	False	45.544132	-34.455868
8	9.65066594	0.43886798	1948.67413	1.54754808	0.03614854	0.3704824	0.31471126	0.55135909	0.36419293	0.1291572	H	21	False	False	False	False	False	False	False	False	False	36.044678	-17.748535
9	9.57186712	0.24851251	1871.20644	1.7407645	0.41662538	0.20200422	0.12607801	-0.2274936	-0.2655372	-0.1770186	H	21	False	False	False	False	False	False	False	False	False	22.526667	-29.994884
10	9.65129894	0.1513891	1369.45773	2.47715704	0.03988407	-0.0216637	0.00721091	-0.2747554	-0.2257595	-0.2033667	H	22	False	False	False	False	False	False	False	False	False	28.284428	-46.640667
11	9.5974207	0.22690312	1762.60733	2.45436658	-0.01384735	-0.030445	-0.0290249	0.20012427	-0.1395469	-0.2134934	H	23	False	False	False	False	False	False	False	False	False	-1.9496368	-52.936028
12	9.27395313	0.06960632	1736.28219	2.18443473	0.02439358	-0.0237361	-0.030316	-0.8776406	-0.6829173	-0.500947	H	23	False	False	False	False	False	False	False	False	False	45.70484	-20.498114
13	9.53834469	0.01505796	1532.49742	1.52459249	0.04477384	-0.0679657	-0.0783795	-0.6219408	-0.4305566	-0.331582	H	23	True	False	False	False	False	False	False	False	False	41.652874	-10.69839
14	10.015581	0.00759823	1827.07564	1.85357929	0.00562961	-0.0529305	-0.0409428	-0.6745631	-0.4116361	-0.3451526	H	24	False	False	False	False	False	False	False	False	False	15.0864315	-9.200474
15	9.14522527	0.17248252	1900.11819	1.68661522	0.82624603	0.50069624	0.22249257	-0.6317193	-0.2673379	-0.1736372	H	24	False	False	False	False	False	False	False	False	False	35.51415	-23.504143
16	9.1714149	-0.2891961	1942.26785	1.76733521	-0.2630754	-0.4182214	-0.2291978	-0.6051788	-0.2729191	-0.1733168	H	24	False	False	False	False	False	True	False	False	False	35.086647	-22.077267
17	9.62154465	0.20500143	1490.58307	2.22489696	0.02180324	0.09650073	0.08814156	0.87986923	0.73544354	0.58445557	H	25	False	False	False	False	False	False	False	False	False	42.91513	-13.328243
18	9.23601852	-0.1326298	1891.78794	1.56224342	0.03100606	-0.0272239	-0.0249314	-0.9408255	-0.577631	-0.3631436	H	25	False	False	False	False	False	False	False	True	False	41.36004	-30.988653
19	9.12881756	0.11234446	2206.95362	1.00858801	0.00030816	-6.12E-05	-7.22E-06	0.00545639	0.02141098	0.00552327	H	25	True	False	False	False	False	False	False	False	False	20.251429	-59.746573
20	9.2959204	-0.125176	1822.67	1.5229196	0.29633123	0.16385714	0.11420808	-0.548561	0.3440676	-0.2188487	H	25	True	False	False	False	False	False	False	False	False	44.36326	-22.203085
21	9.9100645	-0.1249438	2200.0598	1.58930563	0.1087868	-0.0263859	-0.0017723	-0.8711286	-0.534538	-0.1939742	H	27	False	False	False	False	False	False	True	False	False	19.64663	-2.4291084
22	9.63685694	0.04329739	1333.28968	1.63209793	-0.1598486	-0.1457158	-0.0922932	-0.2281525	-0.2920291	-0.1935871	H	27	False	False	False	False	False	True	True	False	False	-5.9635777	-51.227444
23	9.29422457	-0.1415588	1865.56593	1.59993208	-0.0369588	-0.0510069	-0.0627799	-0.77312	-0.5217871	-0.3314062	H	27	True	False	False	False	False	False	False	False	False	44.88403	-14.404646
24	9.4466983	-0.0565235	1548.8748	1.58684793	-0.040572	-0.0781773	-0.0761532	0.25637438	0.08294715	-0.0581436	H	28	False	False	False	False	False	False	True	False	False	43.694935	-13.481288
25	9.86772573	0.02148337	1710.17587	1.86573941	0.08339646	0.02539779	0.04882694	0.57761036	0.96707205	0.96667773	H	29	False	False	False	False	False	False	False	False	False	-7.274528	-55.984974
26	10.0251134	-0.0882456	1856.03339	1.9822702	0.01298564	0.01106014	0.02092347	0.5024589	0.0716915	-0.1438582	H	29	False	False	False	False	False	False	False	False	False	28.428722	-16.47094
27	9.21461064	-0.3231412	1664.39114	1.57716791	0.01224354	0.0075264	0.04034913	0.94424237	0.80682315	0.72535161	H	29	False	False	False	False	False	False	True	False	False	41.726955	-18.358276
28	9.16013769	0.14836907	1601.42569	1.6412938	0.42651234	0.05842841	0.04488505	0.73013871	0.40182242	0.13885073	H	30	False	False	False	False	False	False	False	False	False	41.2001	-32.356304
29	9.80174974	0.20253357	2026.73393	1.72698988	-0.2063181	-0.1932914	-0.1679458	-0.17243	-0.2490311	-0.1753346	H	30	False	False	False	False	False	False	True	False	False	-23.913412	-59.078793
30	9.83257951	0.05130549	1699.39626	1.61346569	0.13614879	0.07450128	0.07931521	0.4446495	0.21168883	0.06401353	H	30	True	False	False	False	False	True	False	False	False	-3.5931506	-68.719154
31	9.88880759	0.15226922	1726.06006	1.61294491	0.18885921	0.05971212	0.04958776	0.28486873	-0.1016129	-0.1869598	H	30	True	False	False	False	False	False	True	False	False	32.391342	-23.016165
32	9.27519381	0.01748547	1887.33448	2.20240525	0.21294463	0.01097064	0.09179338	-0.25692	-0.243346	-0.2102529	H	34	False	False	False	False	False	False	False	False	False	43.86638	-3.8566375
33	10.1024237	-0.3539296	1837.55995	1.99998275	-0.0983338	-0.0404221	-0.0293061	-0.2700099	-0.4073973	-0.2882255	H	35	False	False	False	False	False	False	False	False	False	35.760554	-5.7602686
34	9.57678328	0.12442512	1614.62458	2.07215605	-0.0698632	-0.1274039	-0.1082894	-0.6220778	-0.3030762	-0.1955218	H	35	False	False	False	False	False	False	False	False	False	25.848309	-48.47941
35	9.53589166	0.20334952	1594.29593	1.7676932	0.10724421	0.08910738	0.06517594	0.36392101	-0.042796	-0.1603292	H	35	False	False	False	False	False	False	False	False	False	9.179155	-50.001366
36	9.25297414	0.26063104	1462.1897	1.73598578	-0.3751121	-0.0103524	-0.0631236	0.69097784	0.242779	0.00179909	H	36	True	False	False	False	False	False	False	False	False	41.805363	-36.547287

Figura 11: Exemple conjunt de dades extret pel Matlab

Per extreure'n el fitxer de característiques que podem veure a Fig. 11 s'utilitza un script de Matlab que ha sigut proporcionat. L'entrada d'aquest script són els àudios amb les bufades ja retallats i la sortida és el conjunt de dades en format '.csv'.

2.3. Anàlisi de les dades a SPSSModeler

Per analitzar i predir l'edat/capacitat pulmonar s'ha utilitzat l'eina SPSS Modeler [17]. És una eina enfocada a l'anàlisi de dades i al machine-learning d'una forma visual i intuïtiva. Les característiques extretes pel Matlab en un fitxer '.csv' seran l'entrada que farem servir.

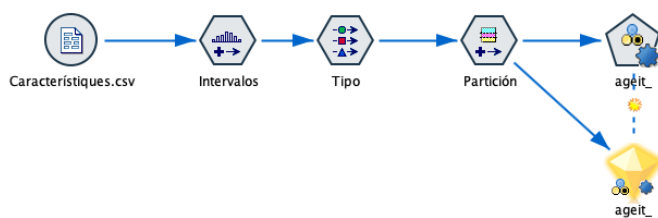


Figura 12: Arxiu SPSS Modeler

Intervalos

Generar Vista previa

Configuración Agrupar valores en intervalos Anotaciones

Campos de intervalos: age

Método de intervalos: Ancho fijo

Intervalos de ancho fijo

Extensión del nombre: it_ Añadir como: ☒ Sufijo ☐ Prefijo

☐ Ancho de intervalo 10,0

☒ N° de intervalos 2

☐ Utilizar los mismos intervalos para todos los campos

Figura 13: Intervals conjunt dades

Partición

Generar Vista previa

Configuración Anotaciones

Campo de partición: Partición

Particiones: ☒ Entrenamiento y comprobación ☐ Entrenamiento, comprobación y validación

Tamaño de partición de entrenamiento: 80 Etiqueta: Entrenamiento Valor = "1_Entrenamiento"

Tamaño de partición de comprobación: 20 Etiqueta: Comprobación Valor = "2_Comprobación"

Tamaño de partición de validación: 0 Etiqueta: Validación Valor = "3_Validación"

Tamaño total: 100%

Valores: ☐ Utilizar valores definidos por el sistema ("1", "2" y "3")

☒ Añadir etiquetas a los valores definidos por el sistema

☐ Utilizar etiquetas como valores

☒ Asignación de partición repetible

Semilla: 1234567 Generar

☐ Utilizar campo exclusivo para asignar particiones:

Aceptar Cancelar Aplicar Restablecer

Figura 14: Entrenament i comprovació

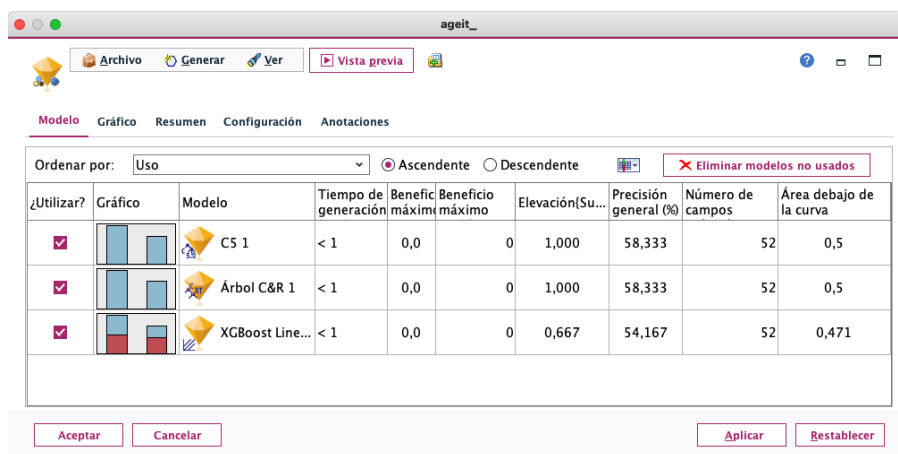


Figura 15: Resultats

El que podem veure a la Figura 12 són tots els passos que SPSS Modeler ha de realitzar per poder fer una predicció de l'edat. El primer pas és l'input de dades el qual es pot veure al nom el fitxer d'entrada (Caracteristiques.csv).

En el següent pas a la Figura 13, anomenat 'Intervalos' configurem els intervals que volem que tingui una columna de dades. Com podem veure a l'exemple de la Figura 16 si separem l'edat en dos intervals, menors de 50 i majors de 50, tindrem 2 grups. Això, ens serveix per agrupar dades similars i obtenir un rendiment de predicció més gran. Com menys intervals tinguem més bona predicció tindrem perquè no hi han dades tan disperses, però els resultats no seran tan precisos perquè només s'estarà separant en 2 grups.

Nom	Edat	Interval Edat
Josep	20	1
Maria	34	1
Anna	41	1
Pol	57	2

Figura 16: Exemple interval



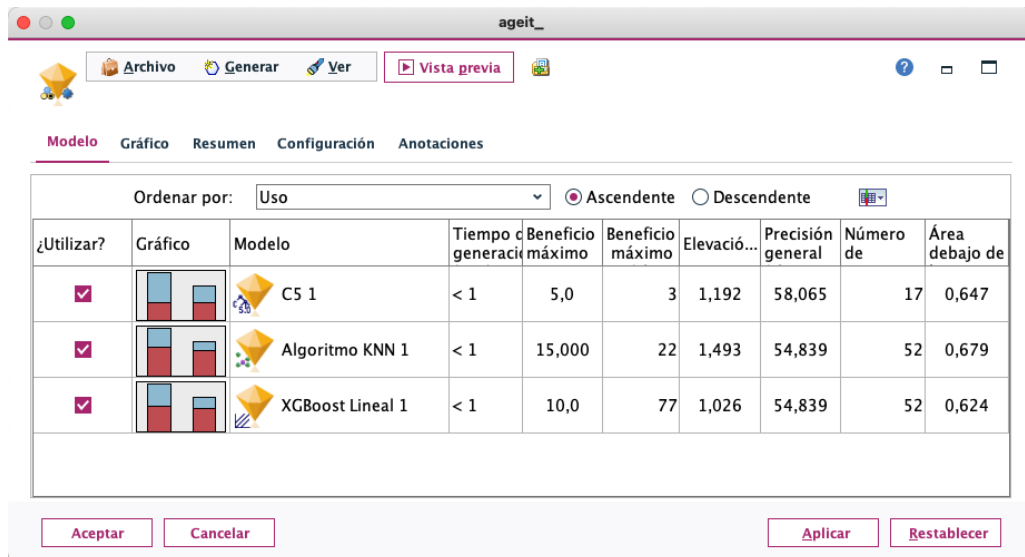
Figura 17: Exemple tipus

El tercer pas, anomenat 'Tipo', configurarem de quin tipus són les variables. És diferent que siguin cadenes de text, nombres decimals, valors booleans, etc. També es configura aquí si les dades ens serveixen d'entrada o de sortida. Si són d'entrada, els algorismes de predicció les utilitzaran per fer la predicció, en canvi si són de sortida significa que seran les dades que s'hauran de predir. En aquest cas el que volem predir és l'edat que hem separat en 2 intervals.

A continuació, el pas anomenat 'Particion' és el que podem veure a la Figura 14. Aquí decidim quin percentatge de dades volem que faci servir el model de machine-learning per entrenar i comprovar. També hi ha altres configuracions com si volem que les separi aleatòriament o ho faci sempre igual.

L'últim pas, anomenat 'ageit_', és afegir els models de machine-learning per realitzar la predicció de l'edat. I un cop executat tot el programa, ens mostrarà en una finestra emergent els resultats de l'execució. Ho podem veure a la Figura 15, que ens diu quin algorisme de machine-learning ha utilitzat, el % d'encert que ha tingut, el nombre de camps que ha fet servir, el temps que ha tardat, etc.

3. Resultats



¿Utilizar?	Gráfico	Modelo	Tiempo de generación	Beneficio máximo	Beneficio máximo	Elevación...	Precisión general	Número de	Área debajo de
<input checked="" type="checkbox"/>		C5 1	< 1	5,0	3	1,192	58,065	17	0,647
<input checked="" type="checkbox"/>		Algoritmo KNN 1	< 1	15,000	22	1,493	54,839	52	0,679
<input checked="" type="checkbox"/>		XGBoost Lineal 1	< 1	10,0	77	1,026	54,839	52	0,624

Figura 18: Resultats anàlisi 2 intervals



¿Utilizar?	Gráfico	Modelo	Tiempo de generación	Beneficio máximo	Beneficio máximo	Elevación...	Precisión general	Número de	Área debajo de
<input checked="" type="checkbox"/>		Á...	< 1	0,0	0	1,833	51,613	52	0,711
<input checked="" type="checkbox"/>		S...	< 1	0,0	0	1,667	45,161	52	0,556
<input checked="" type="checkbox"/>		X...	< 1	10,0	12	1,833	45,161	52	0,633

Figura 19: Resultats anàlisi 3 intervals

Els resultats obtinguts en l'anàlisi del SPSS Modeler, de moment amb 2 intervals tenim un 58% màxim d'encert tal com podem veure en la Figura 18. Altres algorismes han obtingut menys percentatge d'encert, per aquest cas 'C5 1' és el que millors resultats hem obtingut.

En el cas de la Figura 19 s'han utilitzat 3 intervals i obtenim un 51% d'encert amb l'algorisme 'Arbre XGBoost 1'.

Aquests resultats no són dolents perquè prediem l'edat en intervals correctament de més de la meitat, però hi ha molt marge de millora de cara un futur. Aquests resultats no gaire alts es deuen a la poca quantitat de mostres que tenim actualment, unes 150. Un cop tinguem més mostres es podrà predir amb més precisió i amb un percentatge d'encert més alt.

4. Conclusions

Un cop obtingut els primers resultats de predicció, tot i no ser els resultats que volíem, no són dolents. A mesura que hem anat obtenint més mostres hem pogut realitzar petits canvis per millorar parts de codi o maneres d'analitzar. Principalment ens ha costat obtenir mostres i això ens ha causat tenir resultats baixos, però a mesura que puguem tenir més mostres podrem augmentar el percentatge d'encert.

Un incentiu per obtenir mostres i que s'està treballant per tenir és que quan un usuari envia la mostra, ja es pugués fer una predicció. Així l'usuari podria provar com funciona i pot resultar més gratificant participar, ja que tindria resultats del que està enviant.

Referències

- [1] World Health Organization. The top 10 causes of death, 2017.
- [2] Alda Marques, Ana Oliveira, and Cristina Jácome. Computerized Adventitious Respiratory Sounds as Outcome Measures for Respiratory Therapy: A Systematic Review. *Respiratory Care*, 59(5):765–776, 2014.
- [3] B. Bake, P. Larsson, G. Ljungkvist, E. Ljungström, and A. C. Olin. Exhaled particles and small airways. *Respiratory Research*, 20(1):1–14, 2019.
- [4] W A Groves and E T Zellers. Investigation of organic vapor losses to condensed water vapor in Tedlar bags used for exhaled-breath sampling. *American Industrial Hygiene Association journal*, 57(3):257–263, mar 1996.
- [5] V A Goncharova, D T Mamedov, and E K Dotsenko. [Biologically active substance levels in exhaled air from patients with pre-asthma and bronchial asthma]. *Sovetskaia meditsina*, (5):22–24, 1989.
- [6] C T Carpenter, P V Price, and B W Christman. Exhaled breath condensate isoprostanes are elevated in patients with acute lung injury or ARDS. *Chest*, 114(6):1653–1659, dec 1998.
- [7] Mayank Goel, Elliot Saba, Maia Stiber, Eric Whitmire, Josh Fromm, Eric C. Larson, Gaetano Borriello, and Shwetak N. Patel. SpiroCall: Measuring lung function over a phone call. *Conference on Human Factors in Computing Systems - Proceedings*, pages 5675–5685, 2016.
- [8] R J Knudson, R C Slatin, M D Lebowitz, and B Burrows. The maximal expiratory flow-volume curve. Normal standards, variability, and effects of age. *The American review of respiratory disease*, 113(5):587–600, may 1976.
- [9] M. R. Miller, J. Hankinson, Vito Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P.M. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, and J. Wagner. Standardisation of spirometry. *European Respiratory Journal*, 26(2):319–338, 2005.
- [10] Marc Andre Carbonneau, Ghyslain Gagnon, Robert Sabourin, and Jean Dubois. Recognition of blowing sound types for real-time implementation in mobile devices. *2013 IEEE 11th International New Circuits and Systems Conference, NEWCAS 2013*, (January 2016), 2013.
- [11] R. Baken Orlikoff and R. *Clinical measurement of speech and voice*. Singular Pub Group, 2000.
- [12] What is angular? <https://angular.io/guide/what-is-angular>. Accedit: 17-06-2021.

- [13] What is mongodb? <https://www.mongodb.com/what-is-mongodb>. Accedit: 17-06-2021.
- [14] Pàgina web de matlab. <https://www.mathworks.com/products/matlab.html>. Accedit: 17-06-2021.
- [15] Pàgina web de python. <https://www.python.org>. Accedit: 17-06-2021.
- [16] Pàgina web de la llibreria pydub. <https://pypi.org/project/pydub/>. Accedit: 17-06-2021.
- [17] Pàgina web de ibm - spss modeler. <https://www.ibm.com/products/spss-modeler>. Accedit: 20-06-2021.